MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 9: 519–536 (2025)
DOI: 10.3897/mbmg.9.161301

Primer Validation

# The design and testing of taxon-specific mini-barcode markers for metabarcoding of marine zooplankton

Saiesha Harpal[1,2], Johan Groeneveld[1,2], Sandi Willows-Munro[2], Jenny Huggett[3,4], Ashrenee Govender[1]

1 *Oceanographic Research Institute, Durban, South Africa*
2 *School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa*
3 *Oceans and Coastal Research, Department of Forestry, Fisheries and the Environment (DFFE), Cape Town, South Africa*
4 *Department of Biological Sciences, University of Cape Town, Cape Town, South Africa*

*Corresponding author: Ashrenee Govender (agovender@ori.org.za)*

## Abstract

DNA metabarcoding of zooplankton samples is well suited to surveys of marine pelagic ecosystems; however, consistent amplification across zooplankton groups cannot be assumed when using "universal" miniCOI primers. Under-representation of some groups may result from primer bias or inadequate coverage in barcode reference libraries. We designed taxon-specific mini-barcode primers to improve the representation of copepods, euphausiids, chaetognaths, and hydrozoans in metabarcoding outputs of mixed zooplankton samples from the Southwest (SW) Indian Ocean. In silico analyses of downloaded sequences per group identified the most informative mini-barcode region, and barcode gap analyses confirmed that the selected regions could distinguish among species within each group. In vitro analyses of DNA extracts from individual specimens per group showed that the new mini-barcode primers outperformed the standard and universal miniCOI primers by consistently recovering higher Phred quality scores. Metabarcoding of four in situ mixed zooplankton samples collected with plankton tow nets and processed with a combination of taxon-specific and universal primers identified 220 species. Use of the primer cocktails increased the proportionate representation of the target groups by three- to five-fold compared to a previous study. Read counts were dominated by copepods and euphausiids, implying that they had the highest relative biomass in samples. We conclude that a combination of universal and taxon-specific primers in metabarcoding assays will achieve a more comprehensive assessment of biodiversity by enhancing species richness estimates across different groups.

**Key words:** Biodiversity surveys, marine pelagic ecosystems, metabarcoding, mini-barcode primers

OPEN ACCESS

**Academic editor:** Andrew R. Mahon
**Received:** 6 June 2025
**Accepted:** 16 September 2025
**Published:** 23 October 2025

**Citation:** Harpal S, Groeneveld J, Willows-Munro S, Huggett J, Govender A (2025) The design and testing of taxon-specific mini-barcode markers for metabarcoding of marine zooplankton. Metabarcoding and Metagenomics 9: e161301. https://doi.org/10.3897/mbmg.9.161301

**Copyright:** © Saiesha Harpal et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

## Introduction

Metabarcoding is a powerful tool for tracking marine zooplankton diversity as a biologically essential ocean variable (EOV) of pelagic ecosystems (Miloslavich et al. 2018; Muller-Karger et al. 2018). In marine ecological research, species richness derived from metabarcoding enhances community analyses, biodiversity

assessments, ecosystem monitoring, and tracking of climate change impacts on ocean productivity (Huggett et al. 2022; Ershova et al. 2023). Metabarcoding of zooplankton enables the detection of cryptic species, damaged specimens, and multiple life stages that lack clear diagnostic features, such as eggs and larvae (Bucklin et al. 2016; Zhang et al. 2018; Laakmann et al. 2020). However, the full potential of metabarcoding to comprehensively explore zooplankton biodiversity in pelagic ecosystems is not yet realized because of technological and data gaps (Bucklin et al. 2021; Govender et al. 2024).

The mitochondrial cytochrome c oxidase subunit I (COI) region is the preferred marker for barcoding of marine zooplankton (Bucklin et al. 2021). It provides good taxonomic resolution across most zooplankton groups, and extensive barcode reference databases exist that link taxonomically verified species to COI sequences. Open-access data portals for zooplankton barcode data include the MetaZooGene Barcode Atlas and Database (MZGdb), which is linked to NCBI GenBank and the Barcode of Life Database (BOLD) (Bucklin et al. 2021). However, COI does not amplify consistently well across all zooplankton groups. When using "universal" COI primers across distantly related taxonomic groups, amplification bias may occur because of limited primer choice and mismatches (Albaina et al. 2024), the absence of a barcode gap (overlap between inter- and intraspecific genetic distances; Meyer and Paulay 2005), lack of conserved flanking regions, or inherent weaknesses in barcode reference databases (Keck et al. 2023).

Most high-throughput sequencing (HTS) platforms used for metabarcoding (e.g., Illumina MiSeq) have limited read lengths, necessitating the use of mini-barcode primers (shorter gene fragments of 100–350 base pairs [bp]) instead of Folmer's original barcode of ~650 bp (Folmer et al. 1994; Govender et al. 2019). Mini-barcodes offer sufficient variability for species-level resolution while remaining compatible with HTS technologies (Hajibabaei et al. 2006; Hajibabaei and McKenna 2012; Leray et al. 2013; Yeo et al. 2020). For metabarcoding of zooplankton, the universal "miniCOI" barcode (~313 bp; Leray et al. 2013) is commonly used; however, Albaina et al. (2024) reported mismatches in the primer binding sites in some abundant zooplankton taxa and recommended combining complementary taxon-specific and universal primers to enhance detection. Few studies have reported on the design and testing of taxon-specific mini-barcode primers in different zooplankton groups and how combining multiple primers into "primer cocktails" may benefit organismal and eDNA applications (Govender et al. 2019, 2022; Komai et al. 2019).

Several holoplanktonic taxa (copepods, euphausiids, and chaetognaths) as well as hydrozoans (comprising both holoplanktonic and meroplanktonic species) were abundant in visual analyses of zooplankton samples collected with plankton nets off eastern South Africa (SW Indian Ocean) but were under-represented in subsequent metabarcoding outputs (Govender et al. 2023). The sparsity of holoplanktonic species in metabarcoding outputs was partially explained by incomplete barcode reference libraries for these groups in the region (Singh et al. 2021), although Rawoot et al. (2024) added multiple new copepod records from the SW Indian Ocean to BOLD and GenBank. A second explanation, addressed in this study, is primer bias when using standard or universal miniCOI primers across diverse zooplankton groups

(Zhang et al. 2018; Govender et al. 2022). An in-depth literature review by Albaina et al. (2024), which included studies from the Atlantic, Pacific, Indian, Arctic, and Southern oceans, reported misperformance in various zooplankton taxa when using miniCOI (Leray et al. 2013) primer combinations for metabarcoding. Taxon-specific mini-barcode primers substantially improved identification of meroplanktonic species (drifting larvae of lobsters, prawns, shrimps, and crabs) in metabarcoding outputs (Govender et al. 2019, 2022, 2023), compared to lower species identification success rates for abundant holoplanktonic groups amplified with the universal miniCOI primer.

The aims of this study were to: (1) design taxon-specific mini-barcode primers from the COI gene region for Copepoda (copepods), Euphausiacea (krill), Chaetognatha (arrow worms), and Hydrozoa (medusozoan Cnidaria) using in silico methods; (2) test the performance of taxon-specific primers against the standard (~650 bp) COI and the miniCOI primers for zooplankton (Leray et al. 2013); and (3) apply taxon-specific mini-barcode primers, used in combination, in metabarcoding of mixed zooplankton samples collected with plankton tow nets. Metabarcoding outputs (proportionate representation per major taxon) were compared with an earlier study (Govender et al. 2023), which relied on the universal miniCOI primer (Leray et al. 2013) for holoplanktonic species identification.

## Materials and methods

### Taxon-specific primer development

Taxon-specific mini-barcode primers were designed following Govender et al. (2019). For in silico analysis to determine the most variable COI portion, we downloaded sequences of 954 copepod species (1,910 sequences; 110 families; 345 genera), 65 euphausiid species (130 sequences; 2 families; 11 genera), 25 chaetognath species (focusing on Sagittoidea; 50 sequences; 5 families; 13 genera), and 404 hydrozoan species (808 sequences; 83 families; 200 genera) from GenBank and BOLD (www.ncbi.nlm.nih.gov/genbank; http://www.boldsystems.org/; accessed: 1 December 2023; Suppl. material 1). Two individuals per species from different geographical areas were included, where available, to capture within-species variation.

Sequences were aligned separately for each of the four datasets using Clustal X2.1 (Larkin et al. 2007) and manually optimized to ensure homology using BioEdit 7.2.5 (Hall 1999). All possible mini-barcode fragments per dataset were estimated using sliding window analysis (SWAN; Proutski and Holmes 1998) in the Species Identity and Evolution (SPIDER; Brown et al. 2012) package in R (http://www.r-project.org). The “slideAnalyses” function was used to generate windows varying in size from 100 to 300 bp, which were moved along alignments at 10 bp intervals. Two mini-barcode fragments per window length were chosen for further analysis based on: (1) high mean Kimura 2-parameter (K2P) distance; (2) few zero pairwise non-conspecific distances (minimum interspecific distance for each individual); and (3) a high proportion of clades shared between neighbor-joining (NJ) trees derived from the mini-barcode region compared to reference trees constructed from full-length DNA sequence alignments (Govender et al. 2019).

Based on the SWAN analysis, a total of 42 potential mini-barcode fragments were selected for each dataset. Maximum likelihood (ML) analysis was conducted on potential mini-barcode fragments (comparison trees) and full-length sequence alignments (reference trees) for each dataset using Garli 0.951 (Zwickl 2006). The K2P model (Kimura 1980) of sequence evolution was used for all ML analyses, as it is the most often applied by the DNA barcode community and BOLD. Ktreedist (Soria-Carrasco et al. 2007) was used to compare the mini-barcode ML trees to the reference trees. Ktreedist calculates the minimum branch length distance (K-score) from one phylogenetic tree to another. The K-scores (topology and branch length differences) and Robinson–Foulds symmetric differences (topological differences) were also calculated. For both methods, low values imply high similarity between mini-barcode and reference trees. Mini-barcode regions for each dataset were selected based on the lowest K-scores and Robinson–Foulds symmetric differences.

## Evaluation of the target region

DNA barcode gap analyses were conducted to confirm that the selected regions would be able to distinguish between species. The K2P nucleotide substitution model in MEGA 6.0 (Tamura et al. 2013) was used to calculate intra- and interspecific genetic distances. The maximum intraspecific distance value was subtracted from the minimum interspecific distance value to determine the barcoding gap (Meier et al. 2006). The Jeffries–Matusita distance (J–M) statistic was used to test whether the intra- and interspecific genetic distance classes were distinguishable by considering the distance between their means and the distribution of values around the means (Dabboor et al. 2014). The J–M distance is asymptotic to 1.414, with higher values indicating that intra- and interspecific genetic distances are statistically separable (Trigg and Flasse 2001). After confirming the presence of a barcode gap, primers were designed manually for regions flanking the selected mini-barcodes for each of the four groups. When high variability of flanking regions prevented the design of a single primer pair for amplification of all high-level target taxa, multiple primers were designed to form a primer cocktail, specifically for chaetognaths, where two reverse primers were designed. To avoid inadvertent amplification of nuclear mitochondrial insertions (numts), the specificity of each primer was assessed using NCBI's Primer-BLAST tool, confirming that the primers exclusively targeted mitochondrial genes of the intended taxa (https://www.ncbi.nlm.nih.gov/tools/primer-blast/; Suppl. material 1: table S1).

## Primer testing using Sanger sequencing

To evaluate the effectiveness of the mini-barcode primers for copepods (271 bp), euphausiids (310 bp), chaetognaths (283 bp), and hydrozoans (266 bp) (Table 1), amplification success was tested against the universal COI primer set of 658 bp (Folmer et al. 1994) and the miniCOI barcode (Leray et al. 2013) using Sanger sequencing. DNA from 4–5 individual voucher specimens for each of the four groups was extracted using the Zymo Quick-DNA Universal Kit (Zymo Research). An overnight incubation step at 55 °C was added to the standard protocol to ensure complete digestion of tissue in lysis buffer and proteinase K.

**Table 1.** Primers used in this metabarcoding study (first round PCR): each of the COI primer combinations amplifies different fragments of the COI-5P gene region. *Primers were designed for this study.

| Fragment | Primer name | Sequence (5' - 3') | Direction | Reference | Fragment size |
|---|---|---|---|---|---|
| **Copepoda*** | CopeMiniF | GTW ATR CCW ATT TTA ATT GGR GG | F | This study | 271 bp |
| | CopeMiniR | CCT ARA AKA GAA CTA ACT CCT GC | R | | |
| **Euphausiacea*** | EuphausMiniF | CGA GCT GAA YTA GGW CAM CCA GG | F | This study | 310 bp |
| | EuphausMiniR | GCT CCW GCA TGW GCA ATT CCW GC | R | | |
| **Chaetognatha*** | ChaetoMiniF | CCY ACT ATA ATR GGR GGG TTT GG | F | This study | 283 bp - 403 bp |
| | ChaetoMiniR1 | GTA GTR ATR AAA TTW GCW GAT CC | R | | |
| | ChaetoMiniR2 | GTR ATA GCY CCT GCT ART ACA GG | R | | |
| **Hydrozoa*** | HydroMiniF | GCC WGT WTT AAT WGG WGG TTT TGG | F | This study | 266 bp |
| | HydroMiniR | CCC ATW ATW GAW GAA GCW CCW GC | R | | |
| **COI_Leray** | mlCOIintF | GGW ACW GGW TGA ACW GTW TAY CCY CC | F | Leray et al. 2013 | 313 bp–319 bp |
| | HCO2198 | TAA ACT TCA GGG TGA CCA AAA AAT CA | R | Folmer et al. 1994 | |
| **COI_Fish** | mlCOIintF | GGW ACW GGW TGA ACW GTW TAY CCY CC | F | Leray et al. 2013 | 313 bp–319 bp |
| | HCO2198 | TAA ACT TCA GGG TGA CCA AAA AAT CA | R | Folmer et al. 1994 | |
| | FishR2 | ACT TCA GGG TGA CCG AAG AAT CAG AA | R | Ward et al. 2005 | |
| **COI_Lobster** | LobsterMinibarF | GGW GAT GAY CAA ATT TAY AAT G T | F | Govender et al. 2019 | 230 bp |
| | LobsterMinibarR | CCW ACT CCT CTT TCT ACT ATT CC | R | | |

Notes: Illumina adapter target sequences were added to the beginning of each primer set for forward primers (TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN) and reverse primers (GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN). The adapters were used in accordance with the workflow from the Illumina 16S Metagenomics protocol. These adapter targets allow Nextera indexing and Illumina adapter addition through PCR.

PCR reactions (25 µl) contained 20 ng/µl genomic DNA, 12.5 µl OneTaq Quick-Load Master Mix (1X, BioLabs, New England), 0.50 µl forward and reverse primer (10 nmol/L), 7.5 µl sterile nuclease-free water, 1 µl MgCl2 (25 µmol/L), and 1 µl bovine serum albumin (BSA; 1 mg/ml). For primer cocktails, the 0.50 µl primer volume was divided by the number of primers for each forward and reverse reaction; for example, for two reverse primers, 0.25 µl was added for each. Thermal cycling for COI and the newly designed mini-barcode primers comprised an initial denaturation at 94 °C for 2 minutes, 35 cycles of denaturation at 94 °C for 30 seconds, annealing at different primer-specific temperatures for 30 seconds, and extension at 68 °C for 1 minute. The final extension was at 68 °C for 5 minutes. Annealing temperatures were 50 °C for the COI primer, 52 °C for copepods, 56 °C for both euphausiids and hydrozoans, and 53 °C for chaetognaths. A "touchdown" PCR was used for the miniCOI barcode (Leray et al. 2013) to minimize the probability of non-specific amplification. The conditions comprised an initial denaturation at 94 °C for 2 minutes, followed by 16 cycles of denaturation for 10 seconds at 94 °C, annealing for 30 seconds at 62 °C (−1 °C per cycle), extension for 60 seconds at 72 °C, and then an additional 25 cycles at an annealing temperature of 46 °C. The final extension was carried out at 72 °C for 5 minutes. All PCRs included a no-template negative control.

PCR products were visualized on a 2% (w/v) TBE agarose gel containing SafeView™ Classic (Applied Biological Materials Inc., Cat. No. G108). Amplicon size was determined using a 100 bp molecular weight marker (BioLabs, New England). PCR bead clean-up and Sanger sequencing were performed at the Central Analytical Facilities (CAF) at the University of Stellenbosch (South Africa). Sequences were edited using Geneious Prime 2025.0.3, and the percentage of nucleotides with Phred quality scores of at least 30 was calculated for each sequence. Phred scores

indicate the probability that a nucleotide has been correctly identified, serving as a standard measure of sequence quality. A Phred score of 30 corresponds to a 1 in 1000 chance of an incorrect base call and is commonly used as a threshold for high-quality sequence data. The nucleotide BLAST tool (BLASTn) on GenBank and BOLD was used for species identification, with a 97% sequence identity threshold. Sequences generated using the newly designed primers were translated into amino acid sequences using the ExPASy Translate tool (https://web.expasy.org/translate/) and screened for stop codons to verify that nuclear mitochondrial insertions (numts) were not inadvertently amplified. The resulting amino acid sequences were subsequently queried against GenBank using the protein BLAST tool (BLASTp) to confirm that only mitochondrial genes were amplified.

## Application of metabarcoding to bulk zooplankton samples

To evaluate the versatility of the newly designed primers (Table 1) as well as those chosen from literature for Malacostraca (lobsters, crabs, shrimps, prawns, and isopods; Govender et al. 2019, 2022), Actinopterygii (ray-finned fishes; Folmer et al. 1994; Ward et al. 2005; Leray et al. 2013), and the universal primer set (Folmer et al. 1994; Leray et al. 2013), primer testing was carried out on bulk zooplankton samples that were collected with two vertical (200 µm mesh) and two oblique (500 µm mesh) Bongo net tows off eastern South Africa in July 2022 (36.63°S; 23.75°E). For details of previously published primers, refer to Govender et al. (2022).

Whole zooplankton samples were preserved in 97% ethanol during sampling, with the ethanol replaced after 24 hours for long-term storage. In the laboratory, individual samples were homogenized in the 97% ethanol solution for 1 minute using a consumer blender (Milex; 1500 W at 22,000 rpm). Between samples, the blender was washed to remove residual material and rinsed with a 10% bleach solution and 70% ethanol to degrade any remaining DNA. Three subsamples were taken from each homogenate to improve diversity estimates (Lanzén et al. 2017). Each subsample (10 ml of zooplankton homogenate) was centrifuged at 3000 rpm for 1 minute, and excess ethanol was removed; thereafter, 40 mg of tissue was transferred to a sterile tube. The remaining DNA extraction process was carried out as described in Govender et al. (2023).

PCRs were performed in triplicate to reduce the effects of stochasticity, improve accuracy, and minimize bias and amplification errors (Dopheide et al. 2019). PCR reactions (15 µl) contained 20 ng/µl genomic DNA, 7.5 µl Q5 high-fidelity Taq, 0.6 µl forward and reverse primers (10 µM; Table 1), 3.3 µl sterile nuclease-free water, 0.6 µl MgCl2 (25 µM), and 1.2 µl BSA (1 mg/ml). Where primer cocktails were used, the 0.6 µl primer volume was divided by the number of primers for each forward and reverse reaction.

All primers used the same thermal cycling program: initial denaturation at 98 °C for 30 seconds, denaturation at 98 °C for 10 seconds, annealing at different primer-specific temperatures for 30 seconds, and extension at 72 °C for 30 seconds. The final extension step was carried out at 72 °C for 4 minutes. Annealing temperatures were 46 °C for malacostracans, fishes, and universal primers. The annealing temperatures for the newly designed primers were as per the primer testing section above. All PCRs included a no-template negative control. PCR products were visualized on a 2% (w/v) TBE agarose gel containing SafeView™ Classic. Amplicon size was determined using a 100 bp molecular weight marker.

The triplicate PCR products for each of the seven primer sets were pooled and quantified using a Qubit 2.0 Fluorometer (Life Technologies, California, USA), and each of the seven different amplicon pools was further consolidated into a single sample per tow-net haul to create four libraries (one per individual tow) with equimolar concentrations (5 ng/µl). Each library was cleaned using 1.8× Ampure XP purification beads (Beckman Coulter, High Wycombe, UK). Index PCR was performed using the Nextera XT Index Kit (Illumina, San Diego, USA). Thereafter, libraries were cleaned using 0.6× Ampure XP purification beads (Beckman Coulter) and quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit 4.0 instrument (Life Technologies). The four libraries were sequenced on the Illumina MiSeq platform using the MiSeq Nano Reagent Kit v.2 (500 cycles), following the protocols described by Govender et al. (2023) at the KwaZulu-Natal Research and Innovation Platform (KRISP), South Africa.

### Bioinformatic analyses

The DADA2 algorithm (Callahan et al. 2016) implemented in QIIME2 v.2019.10 (Bolyen et al. 2019) was used to conduct quality control checks, chimera removal, filtering, trimming of primers, truncation of forward and reverse reads, and merging of the paired-end reads into amplified sequence variants (ASVs). The first 20 bases were trimmed, and the sequences were truncated at 200 bp for both forward and reverse reads to exclude low-quality regions. The ASVs were queried against BOLD and NCBI GenBank in August 2024 using a 97% sequence identity threshold. ASVs assigned to the same species were merged manually using MS Excel. Species detected by metabarcoding were cross-referenced with occurrence records obtained from online databases such as the World Register of Marine Species (WoRMS; https://www.marinespecies.org), the Ocean Biodiversity Information System (OBIS; https://obis.org), the Global Biodiversity Information Facility (GBIF; https://www.gbif.org), and online literature.

## Results

### Taxon-specific primer development and testing

The most informative region of the COI gene was identified for each of the four groups. The smaller SWAN window sizes (120–140 bp for copepods and hydrozoans, 100–140 bp for euphausiids, and 100–130 bp for chaetognaths) had higher mean K2P distance and lower zero non-conspecific values. Larger window sizes (260–300 bp for copepods, 270–300 bp for euphausiids, 220–300 bp for chaetognaths, and 250–300 bp for hydrozoans) showed better congruence of NJ trees and generated lower K- and R–F scores when compared with reference trees. The intraspecific K2P pairwise distances ranged from 0.00 to 0.14 for copepods, 0.00 to 0.13 for euphausiids, 0.00 to 0.35 for chaetognaths, and 0.00 to 0.16 for hydrozoans. The interspecific distances ranged from 0.00 to 0.67 for copepods, 0.00 to 0.32 for euphausiids, 0.00 to 0.52 for chaetognaths, and 0.00 to 0.79 for hydrozoans. DNA barcode gap analyses based on in silico data showed minimal overlap between intra- and interspecific K2P pairwise distances in all four groups for the selected mini-barcode regions (Fig. 1). All J–M distances were > 1.414, providing statistical support for the existence of a DNA barcode gap in all four datasets.

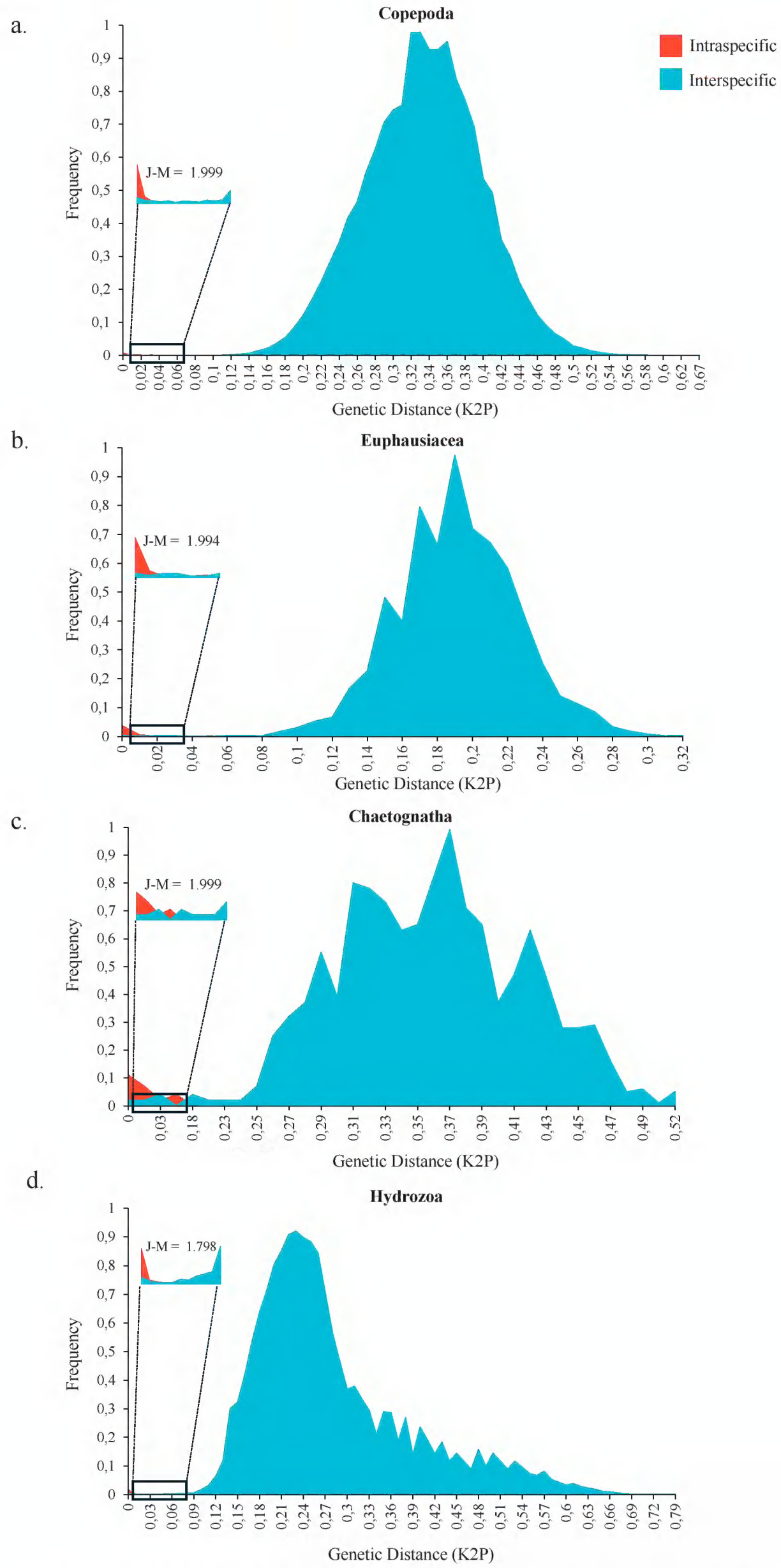


**Figure 1.** Frequency distribution of intra- and interspecific pairwise K2P genetic distances calculated using the selected mini-barcode regions for (**a**) Copepoda, (**b**) Euphausiacea, (**c**) Chaetognatha, and (**d**) Hydrozoa. The frequency data (Copepoda = *n*/120000, Euphausiacea = *n*/1200, Chaetognatha = *n*/100, Hydrozoa = *n*/20000) was normalized to obtain a range between 0 and 1.

Primers were designed within conserved regions flanking the mini-barcode regions (Table 1; Suppl. material 1: table S1). The mini-barcode regions were successfully amplified for all four target taxa using the designed primers, with each primer set tested across the respective samples. BLAST and BOLD search results confirmed that the mini-barcode sequences matched the morphologically identified specimens to at least a family level (93–100% sequence similarity). The amplification success rate of the newly designed mini-barcode primer sets was similar to those of the standard COI primer set and the universal miniCOI primer set (Fig. 2). The sequence quality of the new mini-barcode primers outperformed the standard and universal miniCOI primers by consistently recovering higher Phred quality scores of resulting sequences (Fig. 3). The final primer cocktails used for metabarcoding (Table 1) were based on the amplification and sequencing success of the voucher specimens.

## High-throughput sequencing results and species counts derived from metabarcoding

In total, four zooplankton community libraries were sequenced with Illumina MiSeq. Sequencing was efficient, requiring minimal filtering for both forward and reverse reads during the merging of the paired-end reads for all four zooplankton libraries. For the four libraries, a total of 712,084 read counts were consolidated into 126,449 merged reads, of which 1,456 ASVs were available for analysis across all groups amplified. These ASVs were assigned to a species level at > 97% sequence similarity against reference sequences on BOLD or GenBank and subsequently collapsed into 220 species (Table 2; Suppl. material 1: table S2; Suppl. material 3).

## Species richness and read counts

Species richness was greatest for copepods (80 spp.; 36%), followed by euphausiids (25 spp.; 11%), molluscs (24 spp.; 11%), fishes (23 spp.; 11%), hydrozoans (21 spp.; 10%), decapods (14 spp.; 6%), and chaetognaths (12 spp.; 5%; Fig. 4a). Taxon-specific primers were used for all these groups except for molluscs. Groups for which taxon-specific primers were not used had lower species richness in metabarcoding outputs (<3% per taxon). These taxa included amphipods, echinoderms, ostracods, and scyphozoans. Copepods (41%) and euphausiids (35%) dominated the read counts as a proxy for relative biomass (Fig. 4b), followed by decapods (10%). Chaetognaths (4%) and hydrozoans (5%) had more read counts than any of the other groups for which taxon-specific primers were not used (<2% of read counts per group).

**Table 2.** High-throughput summary statistics for each sample.

| *Library* | *Read count* | *Merged reads* | *Total amplicon sequence variants (ASVs)* | *Species at 97%* |
|---|---|---|---|---|
| ***Oblique 1*** | 155538 | 23331 | 400 | 113 |
| ***Oblique 2*** | 140476 | 24356 | 433 | 115 |
| ***Vertical 1*** | 186622 | 35351 | 522 | 129 |
| ***Vertical 2*** | 229448 | 43411 | 537 | 111 |
| ***Total across nets*** | **712084** | **126449** | **1456** | – |

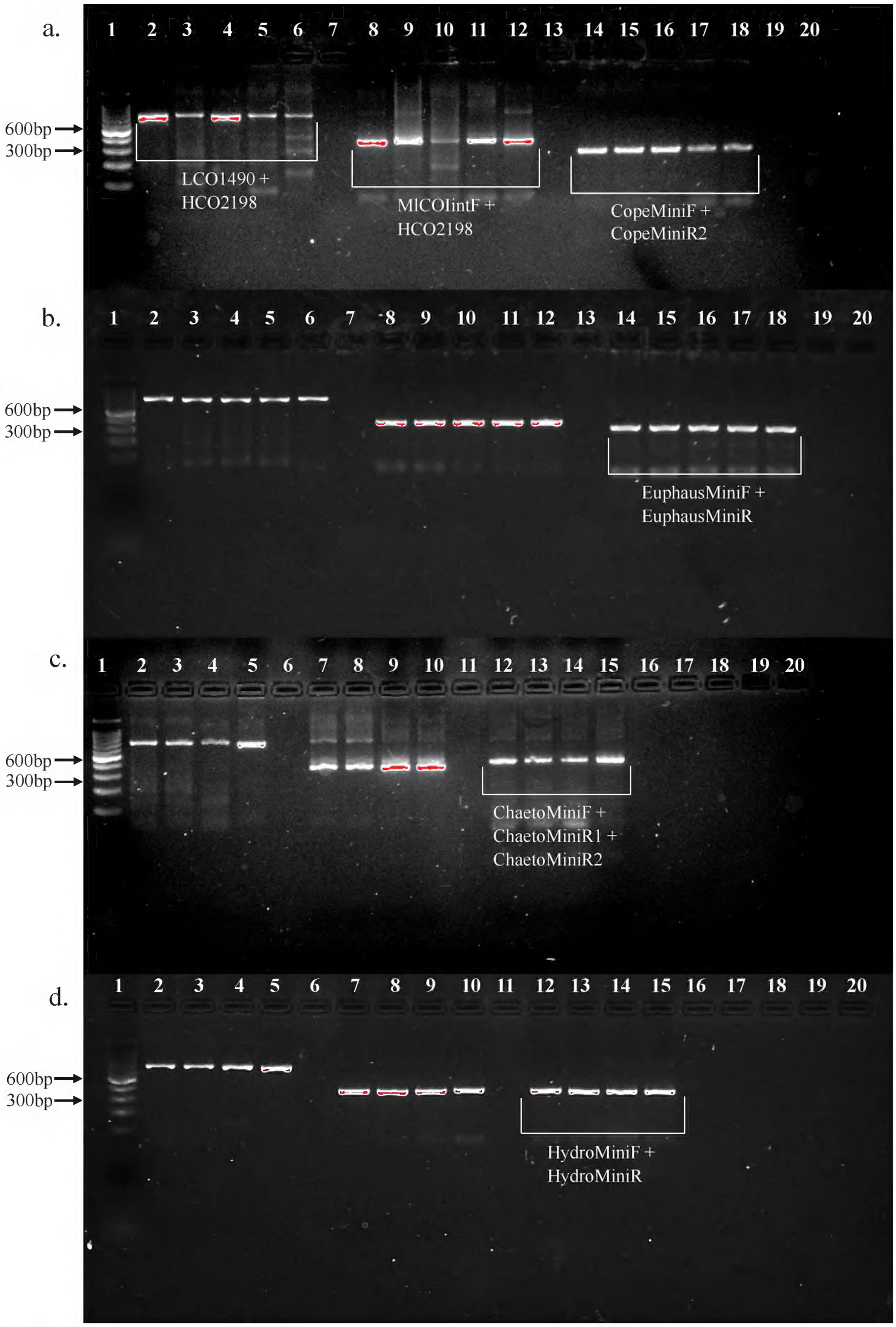


**Figure 2.** Amplification of mini-barcode primers against COI and miniCOI primers for voucher specimens of (**a**) Copepoda, (**b**) Euphausiacea, (**c**) Chaetognatha, and (**d**) Hydrozoa. Lane 1 shows the 100 bp molecular weight marker. Lanes 2–6 (**a, b**) and 2–5 (**c, d**) display PCR products from DNA amplified using universal COI primers. Lanes 8–12 (**a, b**) and 7–10 (**c, d**) show the same samples that were amplified with miniCOI barcode primers. Lanes 14–18 (**a, b**) and 12–15 (**c, d**) show amplification with taxon-specific mini-barcode primers. All PCRs included a no-template negative control in the test gel, although this is not reflected in the gel shown above.

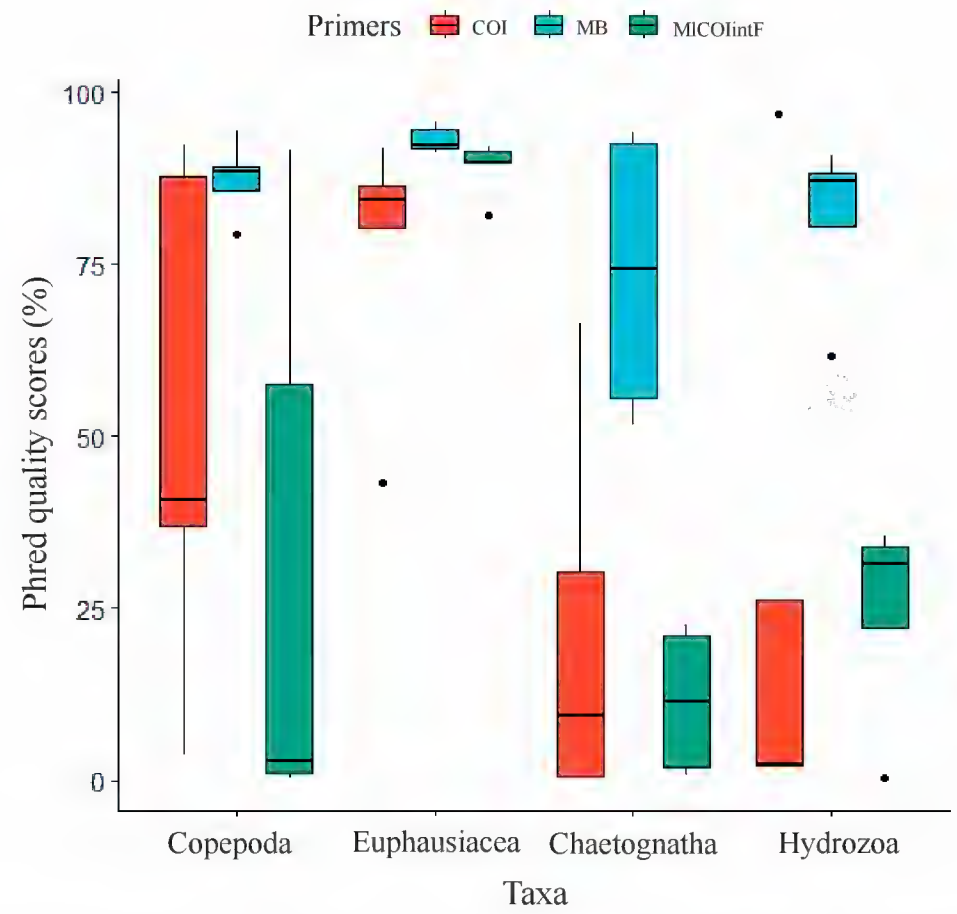


**Figure 3.** Box-and-whisker plot comparing Phred quality scores of COI, mini-barcodes, and universal miniCOI primers for each taxon.

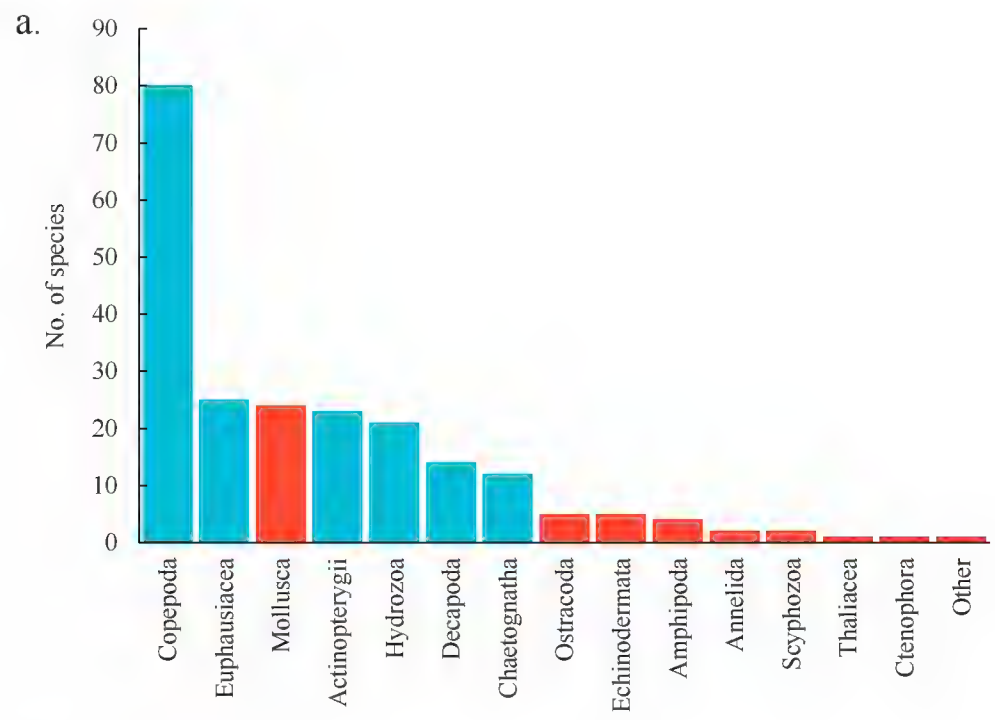


**Figure 4.** Bar graphs showing (a) overall species richness and (b) overall read counts. Blue bars indicate taxa for which specific primers were used.

## Discussion

The integration of DNA-based methods into aquatic monitoring practices is presently on the agenda of several national and international fora (Aylagas et al. 2020; Blancher et al. 2022; Huggett et al. 2022; Ratnarajah et al. 2023). As a contribution to improving biomonitoring capabilities in marine pelagic ecosystems, we designed taxon-specific mini-barcode primers for prominent net-caught zooplankton groups in the SW Indian Ocean. These were then tested on mixed zooplankton samples collected from the SW Indian Ocean. The selected groups were under-represented in recent metabarcoding outputs from the region (Govender et al. 2023) and included both dominant metazoan groups of key ecological and biogeochemical significance (copepods and euphausiids) and understudied soft-bodied or gelatinous groups (chaetognaths and hydrozoans). The new primers for these largely holoplanktonic groups strengthen the capacity for comprehensive biomonitoring of zooplankton in the SW Indian Ocean when used in combination with primers developed previously for meroplanktonic groups important to fisheries (e.g., lobsters, prawns, shrimps, and crabs; Govender et al. 2019, 2022) and for fishes (Ward et al. 2005).

The final selected mini-barcodes ranged between 220 and 300 bp, with smaller window sizes performing statistically worse (higher mean K2P distances and lower congruence in NJ trees). This result is consistent with Yeo et al. (2020), where only very short mini-barcodes (<200 bp) performed poorly for both species- or specimen-level identification, but the performance of moderate-length mini-barcodes (>200 bp) and full-length barcodes was statistically similar. The barcode gap analysis indicated minimal overlap between intra- and interspecific K2P pairwise distances, and the J–M distances were above 1.414 in all four groups. Hence, the existence of a DNA barcode gap could be statistically supported, confirming that the new mini-barcodes can successfully distinguish between species in each group. The new mini-barcode primers consistently amplified the target COI region in each of the four groups, with Phred quality scores showing increased sequence quality compared to the standard COI and miniCOI primers. Our results aligned well with previous studies confirming the utility of taxon-specific mini-barcode markers for overcoming amplification challenges without demonstrable loss of information compared to full-length barcodes (Hajibabaei et al. 2006; Meusnier et al. 2008; Govender et al. 2019; Komai et al. 2019; Yeo et al. 2020; Goncalves et al. 2022).

Primer cocktails used in the in situ metabarcoding analysis identified 220 zooplankton species from four plankton net tows at 97% sequence similarity to published barcode records. Proportionately, copepods comprised 36% of the identified species, an increase from only 10% of species identifiable in a previous metabarcoding study from the same region (Govender et al. 2023), where taxon-specific primers for copepods were not used. In the previous study, primers designed for meroplanktonic taxa combined with the miniCOI primer for metazoans (Leray et al. 2013) identified 271 species from 36 tows, with assemblages dominated by malacostracans (59%) and fishes (17%). In contrast, the use of the newly designed taxon-specific primers in this study yielded proportionate increases of three- to fivefold compared with Govender et al. (2023), elevating species richness for euphausiids (3% to 11%), chaetognaths (1% to 5%), and hydrozoans (3% to 10%), alongside the marked increase in copepods. Based on this comparison, these results demonstrate that primer combinations targeting specific taxa substantially

improve detection rates of ecologically important holoplanktonic groups, which were previously under-represented in metabarcoding outputs.

Read (or sequence) counts were dominated by copepods (41%) and euphausiids (35%), implying that they had the highest relative biomass in samples. These two groups are typically the most abundant (copepods) or are often significantly larger (euphausiids) than most other marine zooplankton, with greater biomass in most oceans (Vereshchaka 2024). Ershova et al. (2021, 2023) found strong correlations between read counts and relative biomass of most zooplankton groups, even though estimates may be affected by primer bias, DNA degradation, and efficiency of DNA extraction on read count accuracy (Bucklin et al. 2016). Therefore, our finding that zooplankton relative biomass is dominated by copepods and euphausiids should be seen as indicative rather than an accurate quantitative estimate.

In conclusion, we designed and tested taxon-specific mini-barcode primers for key holoplanktonic groups (copepods, euphausiids, and chaetognaths) and hydrozoans (comprising both holoplanktonic and meroplanktonic species) to improve their representation in large-scale biodiversity surveys of marine pelagic ecosystems. The new mini-barcode primers consistently amplified the target COI region in each of the four groups, with high sequence quality compared to the standard COI and miniCOI primers. Combinations of taxon-specific primers used in metabarcoding analysis of bulk zooplankton samples increased the proportionate representation of these groups by 3–5 times, at a 97% sequence similarity threshold, compared to a previous study (Govender et al. 2023). We highlight that a combination of universal and taxon-specific metabarcoding assays is essential for achieving a comprehensive assessment of biodiversity by enhancing species richness estimates across different groups in marine ecosystems and increasing the likelihood of detecting rare species.

## Acknowledgements

We gratefully acknowledge postgraduate funding for the first author (SH) from the National Research Foundation (NRF). Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF of South Africa. We thank the Department of Forestry, Fisheries and the Environment (DFFE) and the crew of the *S.A. Agulhas II* for sampling at sea.

## Additional information

### Conflict of interest

The authors have declared that no competing interests exist.

### Ethical statement

No ethical statement was reported.

### Use of AI

No use of AI was reported.

### Funding

This work was supported by the National Research Foundation (NRF).

## Author contributions

Conceptualization: AG, JG. Data curation: AG, SH, JG. Formal analysis: JG, SH, AG. Funding acquisition: AG, JG. Investigation: JG, AG. Methodology: JG, AG. Project administration: AG, JG. Resources: JH, AG. Software: AG. Supervision: SWM, AG, JG. Validation: AG, SH. Visualization: AG. Writing - original draft: SH. Writing - review and editing: AG, SWM, JG, JH.

## Author ORCIDs

Saiesha Harpal https://orcid.org/0000-0002-6628-1453
Johan Groeneveld https://orcid.org/0000-0002-9831-9073
Sandi Willows-Munro https://orcid.org/0000-0003-0572-369X
Jenny Huggett https://orcid.org/0000-0001-9315-8672
Ashrenee Govender https://orcid.org/0000-0002-2860-4610

## Data availability

All raw sequence reads used to perform analyses have been uploaded to NCBI under accession number PRJNA1274484.

## References

Albaina A, Garić R, Yebra L (2024) Know your limits; miniCOI metabarcoding fails with key marine zooplankton taxa. Journal of Plankton Research 46(6): 581–595. https://doi.org/10.1093/plankt/fbae057

Aylagas E, Borja Á, Rodríguez-Ezpeleta N (2020) A step towards the validation of bacteria biotic indices using DNA metabarcoding for benthic monitoring. Molecular Ecology Resources 20(6): 1390–1404. https://doi.org/10.1111/1755-0998.13395

Blancher P, Lefrançois E, Rimet F, Vasselon V, Argillier C, Arle J, Beja P, Boets P, Boughaba J, Chauvin C, Deacon M (2022) A strategy for successful integration of DNA-based methods in aquatic monitoring. Metabarcoding and Metagenomics 6: 215–226. https://doi.org/10.3897/mbmg.6.85652

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS II, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable, and extensible microbiome data science using QIIME 2. Nature Biotechnology 37(8): 852–857. https://doi.org/10.1038/s41587-019-0209-9

Brown SDJ, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) Spider: An R package for the analysis of species identity and evolution,

with particular reference to DNA barcoding. Molecular Ecology Resources 12(3): 562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x

Bucklin A, Lindeque PK, Rodriguez-Ezpeleta N, Albaina A, Lehtiniemi M (2016) Metabarcoding of marine zooplankton: Prospects, progress and pitfalls. Journal of Plankton Research 38(3): 393–400. https://doi.org/10.1093/plankt/fbw023

Bucklin A, Peijnenburg KTCA, Kosobokova KN, O'Brien TD, Blanco-Bercial L, Cornils A, Falkenhaug T, Hopcroft RR, Hosia A, Laakmann S, Li C, Martell L, Questel JM, Wall-Palmer D, Wang M, Wiebe PH, Weydmann-Zwolicka A (2021) Toward a global reference database of COI barcodes for marine zooplankton. Marine Biology 168(6): 78. https://doi.org/10.1007/s00227-021-03887-y

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Dabboor M, Howell S, Shokr M, Yackel J (2014) The Jeffries Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data. International Journal of Remote Sensing 35(19): 6859–6873.

Dopheide A, Xie D, Buckley TR, Drummond AJ, Newcomb R (2019) Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution 10(1): 120–133. https://doi.org/10.1111/2041-210X.13086

Ershova EA, Wangensteen OS, Descoteaux R, Barth-Jensen C, Præbel K (2021) Metabarcoding as a quantitative tool for estimating biodiversity and relative biomass of marine zooplankton. ICES Journal of Marine Science 78(9): 3342–3355. https://doi.org/10.1093/icesjms/fsab135

Ershova EA, Wangensteen OS, Falkenhaug T (2023) Mock samples resolve biases in diversity estimates and quantitative interpretation of zooplankton metabarcoding data. Marine Biodiversity 53(5): 66. https://doi.org/10.1007/s12526-023-01372-x

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3: 294–299.

Goncalves LT, Françoso E, Deprá M (2022) Shorter, better, faster, stronger? Comparing the identification performance of full-length and mini-DNA barcodes for apid bees (Hymenoptera: Apidae). Apidologie 53(5): 55. https://doi.org/10.1007/s13592-022-00958-x

Govender A, Groeneveld JC, Singh SP, Willows-Munro S (2019) The design and testing of mini-barcode markers in marine lobsters. PLoS ONE 14(1): e0210492. https://doi.org/10.1371/journal.pone.0210492

Govender A, Singh S, Groeneveld J, Pillay S, Willows-Munro S (2022) Experimental validation of taxon-specific mini-barcode primers for metabarcoding of zooplankton. Ecological Applications : A Publication of the Ecological Society of America 32(1): e02469. https://doi.org/10.1002/eap.2469

Govender A, Singh S, Groeneveld J, Pillay S, Willows-Munro S (2023) Metabarcoding analysis of marine zooplankton confirms the ecological role of a sheltered bight along an exposed continental shelf. Molecular Ecology 32(23): 6210–6222. https://doi.org/10.1111/mec.16567

Govender A, Willows-Munro S, Singh SP, Groeneveld JC (2024) Net type, tow duration and day/night sampling effects on the composition of marine zooplankton derived from metabarcoding. Metabarcoding and Metagenomics 8: e119614. https://doi.org/10.3897/mbmg.8.119614

Hajibabaei M, McKenna C (2012) DNA Mini-barcodes. Methods in Molecular Biology (Clifton, N.J.) 858: 339–353. https://doi.org/10.1007/978-1-61779-591-6_15

Hajibabaei M, Smith MA, Janzen D, Rodriguez J, Whitfield J, Hebert P (2006) A minimalist barcode can identify a specimen whose DNA is degraded: BARCODING. Molecular Ecology Notes 6(4): 959–964. https://doi.org/10.1111/j.1471-8286.2006.01470.x

Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

Huggett JA, Groeneveld JC, Singh SP, Willows-Munro S, Govender A, Cedras R, Deyzel SH (2022) Metabarcoding of zooplankton to derive indicators of pelagic ecosystem status. South African Journal of Science 118(11/12): 1–4. https://doi.org/10.17159/sajs.2022/12977

Keck F, Couton M, Altermatt F (2023) Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources 23(4): 742–755. https://doi.org/10.1111/1755-0998.13746

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16(2): 111–120. https://doi.org/10.1007/BF01731581

Komai T, Gotoh RO, Sado T, Miya M (2019) Development of a new set of PCR primers for eDNA metabarcoding decapod crustaceans. Metabarcoding and Metagenomics 3: e33835. https://doi.org/10.3897/mbmg.3.33835

Laakmann S, Blanco-Bercial L, Cornils A (2020) The crossover from microscopy to genes in marine diversity: From species to assemblages in marine pelagic copepods. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 375(1814): 20190446. https://doi.org/10.1098/rstb.2019.0446

Lanzén A, Lekang K, Jonassen I, Thompson EM, Troedsson C (2017) Correction: DNA extraction replicates improve diversity and compositional dissimilarity in metabarcoding of eukaryotes in marine sediments. PLoS ONE 13(1): e0192337. https://doi.org/10.1371/journal.pone.0192337

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) 23(21): 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10: 34. https://doi.org/10.1186/1742-9994-10-34

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA Barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Systematic Biology 55(5): 715–728. https://doi.org/10.1080/10635150600969864

Meusnier I, Singer GA, Landry JF, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9: 214. https://doi.org/10.1186/1471-2164-9-214

Meyer CP, Paulay G (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLoS Biology 3(12): e422. https://doi.org/10.1371/journal.pbio.0030422

Miloslavich P, Bax NJ, Simmons SE, Klein E, Appeltans W, Aburto-Oropeza O, Andersen Garcia M, Batten SD, Benedetti-Cecchi L, Checkley Jr DM, Chiba S, Duffy JE, Dunn DC, Fischer A, Gunn J, Kudela R, Marsac F, Muller-Karger FE, Obura D, Shin YJ (2018) Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. Global Change Biology 24(6): 2416–2433. https://doi.org/10.1111/gcb.14108

Muller-Karger FE, Hestir E, Ade C, Turpie K, Roberts DA, Siegel D, Miller RJ, Humm D, Izenberg N, Keller M, Morgan F, Frouin R, Dekker AG, Gardner R, Goodman J, Schaeffer B, Franz BA, Pahlevan N, Mannino AG, Concha JA, Ackleson SG, Cavanaugh KC, Romanou A, Tzortziou M, Boss ES, Pavlick R, Freeman A, Rousseaux CS, Dunne J, Long MC, Klein E, McKinley GA, Goes J, Letelier R, Kavanaugh M, Roffer M, Bracher A, Arrigo KR, Dierssen H, Zhang X, Davis FW, Best B, Guralnick R, Moisan J, Sosik HM, Kudela R, Mouw CB, Barnard AH, Palacios S, Roesler C, Drakou EG, Appeltans W, Jetz W (2018) Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems. Ecological Applications : A Publication of the Ecological Society of America 28(3): 749–760. https://doi.org/10.1002/eap.1682

Proutski V, Holmes E (1998) SWAN: Sliding window analysis of nucleotide sequence variability. Bioinformatics (Oxford, England) 14(5): 467–468. https://doi.org/10.1093/bioinformatics/14.5.467

Ratnarajah L, Abu-Alhaija R, Atkinson A, Batten S, Bax NJ, Bernard KS, Canonico G, Cornils A, Everett JD, Grigoratou M, Ishak NHA, Johns D, Lombard F, Muxagata E, Ostle C, Pitois S, Richardson AJ, Schmidt K, Stemmann L, Swadling KM, Yang G, Yebra L (2023) Monitoring and modelling marine zooplankton in a changing climate. Nature Communications 14(1): 564. https://doi.org/10.1038/s41467-023-36241-5

Rawoot A, Govender A, Groeneveld J, Willows-Munro S, Cedras R (2024) Strengthening the DNA barcode reference library for marine copepods in South Africa. African Journal of Marine Science 46(4): 1–9. https://doi.org/10.2989/1814232X.2024.2418573

Singh SP, Groeneveld JC, Huggett J, Naidoo D, Cedras R, Willows-Munro S (2021) Metabarcoding of marine zooplankton in South Africa. African Journal of Marine Science 43(2): 147–159. https://doi.org/10.2989/1814232X.2021.1919759

Soria-Carrasco V, Talavera G, Igea J, Castresana J (2007) The K tree score: Quantification of differences in the relative branch length and topology of phylogenetic trees. Bioinformatics (Oxford, England) 23(21): 2954–2956. https://doi.org/10.1093/bioinformatics/btm466

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA 6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30(6): 2725–2729. https://doi.org/10.1093/molbev/mst197

Trigg S, Flasse S (2001) An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. International Journal of Remote Sensing 22(13): 2641–2647. https://doi.org/10.1080/01431160110053185

Vereshchaka A (2024) Navigating the Zooplankton Realm: Oceans of Diversity Beneath the Sea Surface. Diversity 16(12): 717. https://doi.org/10.3390/d16120717

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD (2005) DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B. Biological 360(1462): 1847–1857. https://doi.org/10.1098/rstb.2005.1716

Yeo D, Srivathsan A, Meier R (2020) Longer is Not Always Better: Optimizing Barcode Length for Large-Scale Species Discovery and Identification. Systematic Biology 69(5): 999–1015. https://doi.org/10.1093/sysbio/syaa014

Zhang GK, Chain FJJ, Abbott CL, Cristescu ME (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications 11(10): 1901–1914. https://doi.org/10.1111/eva.12694

Zwickl D (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD Thesis, The University of Texas, Austin.

## Supplementary material 1

### Additional tables

Authors: Saiesha Harpal, Johan Groeneveld, Sandi Willows-Munro, Jenny Huggett, Ashrenee Govender

Data type: docx

Explanation note: **table S1.** Table showing primer design parameters. **table S2.** List of the 220 species detected by metabarcoding, with percentage sequence similarity to barcode records on BOLD and GenBank.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.161301.suppl1

## Supplementary material 2

### Spreadsheet 1

Authors: Saiesha Harpal, Johan Groeneveld, Sandi Willows-Munro, Jenny Huggett, Ashrenee Govender

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.161301.suppl2

## Supplementary material 3

### Spreadsheet 2

Authors: Saiesha Harpal, Johan Groeneveld, Sandi Willows-Munro, Jenny Huggett, Ashrenee Govender

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.161301.suppl3